DOCUMENT RESUME

ED 195 578 TM 810 032

| | |
|---|---|
| AUTHOR | Hensley, Susan E.: Nelson, Carnot E. |
| TITLE | The Reviewing Processes in Social Science Publications: A Review of Research. ERIC/TM Report 77. |
| INSTITUTION | ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J. |
| SPONS AGENCY | National Inst. of Education (ED), Washington, D.C. |
| PUB DATE | Dec 80 |
| CONTRACT | 400-78-0003 |
| NOTE | 38p. |
| AVAILABLE FROM | ERIC/TM, Educational Testing Service, Princeton, NJ 08541 ($4.00 each). |
| EDRS PRICE | MF01/PC02 Plus Postage. |
| DESCRIPTORS | Editing: *Evaluation Criteria: *Evaluation Methods: *Publications: Scholarly Journals: *Social Sciences: Technical Writing |
| IDENTIFIERS | *Manuscripts |

ABSTRACT

Because of the increase in the number of manuscripts submitted to journals, editors and publishers must either publish more manuscripts or increase the proportion of manuscripts rejected. Rejection decisions are usually informed by informal peer evaluation, editorial processes, and/or citation analysis. Unfortunately, no standard criteria for evaluation have been developed, and inter-referee agreement on the attributes of journal articles is generally low. The research on these issues is reviewed. (Author)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED195578

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

RECEIVED
ERIC/TM

ERIC/TM Report 77

# THE REVIEWING PROCESSES IN SOCIAL SCIENCE PUBLICATIONS: A REVIEW OF RESEARCH

Susan E. Hensley and Carnot E. Nelson

University of South Florida

December 1980

ERIC Clearinghouse on Tests, Measurement, and Evaluation

Educational Testing Service, Princeton, N.J. 08541

TM 810 032

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U. S. Department of Education. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the National Institute of Education.

ERIC Clearinghouse on Tests, Measurement,
& Evaluation
Educational Testing Service
Princeton, NJ 08541

## INTRODUCTION

An important form for the dissemination of scientific information is publication in archival journals. Publication of information in journals serves a number of functions. First, it provides an archival repository for information which the scientific community, through the process of peer review, has deemed of sufficient quality to merit publication. Second, it is the vehicle used by scientists in claiming the right of discovery, at least in the physical sciences (Brittain 1970). Third, publications are used extensively by university administrators to evaluate faculty members. Finally, it is through the publication of his/her work in high-quality journals that the scientist obtains prestige and recognition from the scientific community.

As a number of authors have pointed out (e.g., Storer 1966; Crane 1969), science is a social system in which interactive communication is the most salient aspect. Thus, the scientist is both a producer and user of scientific information. As a user of scientific information, the scientist is in constant need of up-to-date, quality information. Although the nature and source of these information needs differ depending on the scientist's stage in the research process (e.g., perception or definition of problem, selection of data gathering techniques, placing the data in proper context with existing data, etc.), journal articles are important sources of information in all stages of research. For example, Garvey, Lin, and Tomita (1972) found that in the early stages of research, journals provided the most information followed by local colleagues. During the intermediate stage, this order was reversed. In the final stages of research, journals were again the most important source of information.

To the extent that journals either cannot or do not provide current information, scientists develop new forms of communication. For example, scientists, because of their need to learn of the latest developments in their fields and slowness of the journal publication process, have developed informal communication networks. Examples of such informal networks include colloquia, the exchange of preprints, and the attendance at and requests for professional meeting presentations.

Communication in science, especially publication in archival journals, is governed by very strong social norms. One of these norms is the process of routine refereeing of manuscripts by peers. This process serves a number of functions. First, it eliminates "crankiness, irrelevance, and gross incompetence" (Ziman 1970). Second, it minimizes editorial arbitrariness and, third, it provides a stamp of approval by the scientific community as to the quality of the work. As Zuckerman and Merton (1971) pointed out, peer review has historical roots that extend back to the beginnings of the first scholarly journals. Currently, peer review is almost universally used by American and British Commonwealth journals (Manheim 1973). The strength of the norm of peer review is shown in the bitterness that has surrounded attempts in physics and psychology to institute formal preprint groups (i.e., formal distribution of nonrefereed manuscripts).

Although scientists tend to believe that the peer review system is fair and impartial, there is abundant evidence that this may not be the case, especially in the social sciences. For example, Pfeffer, Leong, and Strehl (1977) found that particularism (i.e., the relationships between institutional representation on editorial boards and institutional

contributions to major journals, controlling for institutional size and quality) was greatest in political science, followed by sociology. It was least apparent in chemistry. In addition, Yoels (1974) found that social science editors are more likely than physical science editors to employ particularistic criteria when selecting editorial board members. He found that in the social sciences Columbia and Harvard graduates were more likely to select fellow graduates than were physical science graduates. Lindsey (1978), as part of his extensive study of the publication system in the social sciences, found that while the editorial boards of psychology and sociology journals are staffed by distinguished scientists this is not true of social work.

Publication in journals provides the author with claims to discoveries discussed in the articles which in turn influence his/her relative status in the scientific community. More importantly, for the young scientist, publication in refereed journals is commonly used as a performance measure by university committees and administrators in making promotion and tenure recommendations and decisions. Not only are administators using the number of articles as a measure of performance, but they are also using the number of citations to a faculty member's work as a measure of the quality and impact of the research. Because of the depressed market in academics and the difficulty faced by junior faculty in obtaining promotions and tenure, there is greater pressure on junior faculty to publish. This outside pressure is in addition to the scientists' self-induced need to publish. The increase in pressure to publish is one of the factors that has led to what has been termed an "information explosion." Other causes for the information explosion include the growth in the

absolute number of scientists and the emphasis on analytical pieces by "normal science" rather than theoretical works. Whether this information explosion is a myth, because scientific information has been growing exponentially for centuries (Price 1964; Baker 1970), or not, critical thresholds have been reached which are affecting the scientists' behavior. As Manheim (1973) pointed out, many libraries have abandoned archival commitments, journals and monographs written in foreign languages are less readily available, and scientists are reading a smaller proportion of the literature even within their own specialities, all of which are indications that a critical threshold has been reached.

Because of the increase in the number of manuscripts submitted, journal editors and publishers are faced with a real dilemma. They must either accept and publish more manuscripts, thereby increasing the cost of their publications, or they must increase the proportion of manuscripts rejected, thereby possibly not publishing material that should be published. However, much of the increase in submissions may be due to manuscripts that are inappropriate for the specific readership. Also, editors are not satisfied with the quantity of high-quality manuscripts. Even so, neither of the above-mentioned alternatives appears viable.

With the cost of publishing journals increasing rapidly and the amount of funds available to libraries for purchasing these periodicals remaining stable or decreasing, publishers are caught in another dilemma. (Granted there are exceptions to the statement regarding libraries; some libraries are increasing the amount of funds for periodicals but typically at the expense of book orders.) If publishers increase the size of their journals, they may well price them out of the market. Although size is

only one factor related to pricing, social science journals are limited in the number of pages that they may publish.

One solution to this problem that is frequently used in the physical sciences is the use of page charges whereby the authors are requested to bear a proportion of the costs of publishing their work. However, if grant and institutional funds for paying page charges do not increase at the same rate as publication costs, the publishers are still caught in the same dilemma. According to Garvey, Lin, and Nelson (1970) and Lindsey (1978), the use of page charges is probably one reason that rejection rates in physical science journals are so much less than in social science journals. There are, in addition, other reasons for the high rejection rates in the social sciences (e.g., the lack of an adequate paradigm and the particularistic decisions that are made).

Rejection rates of 80% to 90% are not unusual in the social sciences compared to the 24% in the physical sciences reported by Zuckerman and Merton (1971). These extremely high rejection rates appear to have had three major ramifications. First, social scientists whose manuscripts are rejected by one journal typically submit their manuscripts to different journals until they are finally published. Second, high rejection rates by the major journals in a field can, in conjunction with other factors, lead to the development of new journals. These factors include paradigmatic thrusts deemed inappropriate by the leadership in the field and the need to reach special audiences. These rejected manuscripts are not typically of poor quality; rather, they may not be in the mainstream of the field or may not be significant "breakthroughs" in the mind of the editor or referees. Third, a number of articles have questioned the impartiality,

reliability, and validity of the reviewing process (e.g., Yoels 1974; Scott 1974; Pfeffer, Leong, and Strehl 1977; and Mahoney 1977). The major criticism of the review process in the social sciences centers around the use of particularistic criteria instead of "universalistic" ones. Thus, a number of journals have gone to blind reviewing where the author is unknown to the referee. Whether this technique is useful is still an open question.

This increased concern with the refereeing process has led to a number of studies that have tried to examine various aspects of this process. In this review we will first discuss the methods used by the various scientific communities in evaluating manuscripts, and then we will review the research findings related to the criteria used in the evaluation process, errors in the evaluation process, and the reliability and validity of referee ratings and other criteria.

## METHODS OF EVALUATION

The methods for evaluating research journal articles vary a great deal in terms of their rigor and objectivity. Most articles, however, are subjected to a number of evaluation procedures both before and after publication so that the weaknesses of one approach are counterbalanced by the strengths of others. Informal peer evaluation and the editorial process occur before publication, while citations occur after publication.

Informal Peer Evaluation. The submission of a manuscript to journals typically represents the last step in a process by which the scientist disseminates his/her work. Prior to manuscript submission, the scientist has typically disseminated his/her work through a number of informal

communication channels. One of the most important results of this dissemination is that it allows for the provision of feedback so that the manuscript can be modified and revised prior to its submission.

According to Garvey and his colleagues (Garvey, Lin, and Tomita 1972), four-fifths of the 3,676 authors they surveyed, from eight different physical science, social science, and engineering disciplines, made some sort of prepublication report of the main content of their published article. The most frequently used forums for oral reports were colloquia within the author's own institution (29% of the authors) and meetings of a national society (24%). Although national meeting presentations are reviewed, the process is much less rigorous than that for journal articles. Technical reports (produced by 21% of the authors) and theses or dissertations (produced by 19%) were the most frequently used written channels of communication.

These dissemination activities did in fact provide useful feedback to the authors. About one-half of the authors who made prepublication reports indicated that they received feedback from these reports that led them to modify the main content of the work prior to its submission. About 40% of the authors making prepublication reports stated that they made major changes in their work (e.g., clarification or redefinition, new or further explication of theory, incorporation of another researcher's findings), while 25% modified the style or organization of the manuscript.

Nelson (1972), in his study of educational researchers, found that 70% of his sample of authors made some type of prepublication report and 45% of these authors modified their work as a result of such dissemination activities. Changes in content accounted for 60% of the modifications, while changes in style accounted for 40% of the modifications.

Another major form of prepublication dissemination is the distribution of preprints (copies of the manuscripts). In the Garvey, Lin, and Tomita study (1972), one-third of the authors distributed preprints prior to the submission of their manuscripts, one-fifth between submission and notification of acceptance, and one-sixth after receiving notification of acceptance. The dissemination of preprints prior to submission provides one more opportunity for the authors to receive informal feedback from their peers. About two-fifths of the authors distributing preprints reported receiving feedback that led them to modify their manuscripts. Two-thirds of the authors made stylistic changes, while three-fifths made substantive modifications. Obviously, some authors made both types of changes.

In the study of educational researchers, Nelson (1972) reported that one-quarter of the authors distributed preprints prior to journal submission and that 56% of those that did received useful feedback that led them to modify their manuscripts. Half of these modifications involved stylistic changes and half involved changes in content.

Although informal peer evaluation is not normally considered part of the review process, it would appear that it is heavily used by authors. This type of evaluation allows the author to receive feedback and suggestions concerning his/her work from colleagues and associates prior to submitting the manuscript for editorial review. Based on the available data, it is clear that authors frequently make major changes in their manuscripts as a result of presenting their work to colleagues.

Editorial Processes. Upon submission to a journal, the author's manuscript is subject to a review process. For most journals in the

social sciences, this involves review by both the editor and peers, while for some multidisciplinary European journals, such as Doklady Akademi and Comptes Rendus, much greater control is held by editors (Manheim 1973). However, editors in the social sciences do reject manuscripts if they are obviously inappropriate for the readership of the journal or of extremely poor quality. Beyer (1978) found that 12.5% of the manuscripts submitted to sociology journals and 25.95% of those submitted to political science journals were rejected by the editor without review by referees. (For a discussion of the editor's role, see Balaban 1978.)

Existing editorial review processes, especially in the social sciences, have been subjected to criticism, and the advantages and disadvantages of specific methods of editing have been discussed in recent studies. In 1966, Newman presented criticisms and suggestions for improving the review process for American Psychological Association journals. Among his criticisms were high rejection rates and the lack of validity of evaluations. As a follow-up to these comments, Brackbill and Korten (1970) conducted a survey of psychologists' attitudes toward journal reviewing practices and suggestions for improvement.

The results indicated that the respondents were concerned with the problem of publication lag, time between a manuscript's submission and publication. They agreed something should be done to shorten the review process, but they were not in agreement as to how this should be handled. A second concern dealt with the reviewers themselves. Respondents expressed some skepticism about the knowledge, values, and goals of the reviewers. Finally, respondents expressed a desire for multiple reviewers which would include a review by the editor and peers. According to the

authors, this indicates that "clearly, most authors do seek and value critical appraisal of their work" (p. 940).

In his article on interreferee agreement on manuscript evaluation, Scott (1974) pointed out that high rejection rates in APA journals were still a problem eight years after Newman's criticism. He cited rejection rates for most APA journals of over 70% and proposed two possible solutions. The first solution would be to increase page allocations to accommodate more manuscripts. The problem with this solution is that it increases the burden of the reader who is already saturated with information. The second solution is to "increase the rejection rate until the number of submitted manuscripts declines" (p. 698).

Garvey, Lin, and Nelson (1970) discussed interdisciplinary differences in terms of publication lag time and journal rejection rates. According to the Garvey et al. report (1970), within a year after presenting their work at a national meeting, only one-third of social scientists had their submitted manuscripts published compared to 60% of the physical scientists. This differential publication rate is attributed to high rejection rates in social science journals and to the fact that most social science journals do not use page charges. Under such a system, publication costs are shared by the journal and the author, which means page allocations are not a strict function of annual publication budgets. Therefore, more articles are printed each year. Lindsey (1978) also discussed the impact of page charges on the publication systems of social and physical sciences. Turning to more specific aspects of the manuscripts' review process, one factor that has been the subject of discussion is whether the reviewer should be anonymous or identified. Manheim (1973) reviewed arguments for

both. He listed five factors to support disclosing the identify of the reviewer:

1. With authority should go responsibility. This implies that a man who must make critical judgments about a manuscript should not be allowed to hide behind anonymity, but should be willing to stand up and reveal himself.
2. Harsh criticism may be more acceptable from an authority whose work is respected than from "an unidentified judge out of Kafka."
3. Authors frequently profit by correspondence with named referees, and even co-authorship may be developed by this type of contact.
4. Concealment of a reviewer's identity may be difficult and may promote a distasteful psychological atmosphere of secrecy, incidivism and privilege.
5. Partisan judgments on controversial questions or improper use of privileged information are less tempting if the reviewers' identities are disclosed. (p. 534)

Conversely, Manheim also discussed arguments in favor of reviewer anonymity. He cited the need to separate the reviewer from an interpersonal relationship with the author that may affect the reviewer's objectivity. He offered three additional points that support reviewer anonymity. The first is that harsh criticism may be more acceptable from an authority whose work is respected than from someone not considered an authority, independent of the validity of the evaluation. That is, the author's reaction to criticism will not be a function of the accuracy of the evaluation as much as the reputation of the reviewer. The second point concerns the fact that reviewers are subject to editorial

constraints. If it is apparent that they are consistently prejudiced, they will not be included as reviewers in the future. The third point is that while author-reviewer interaction may result in constructive feedback, it may also have the undesirable side effect of creating in the author an expectation that flaws in manuscripts will be taken care of during the editorial process.

> Moreover, to expect that referees--often the best and busiest people in a profession--should routinely fill in gaps or supply missing expertise alters their role from that of a referee or arbiter to that of a co-worker or even a schoolmaster. Encouraging this trend will do nothing to curb sloppiness or slacken the onrush of papers that threaten to engulf us all. (P. 536)

A related issue deals with whether the author's name and affiliation should be known to the reviewer. In the survey conducted by Brackbill and Korten (1970), the authors found fairly high agreement among respondents to the item, "An author's name and institutional affiliation should be deleted from a manuscript before it is reviewed." Such a procedure is designed to reduce the possibility that the author's prestige or institutional affiliation will influence the referee. However, many reviewers may still know whose work is being considered based on other professional contacts and the size of the speciality.

In a study similar to the one conducted by Brackbill and Korten (1970), Silverman and Collins (1975) examined the attitudes of members of the American Association for Higher Education and the editors of journals of higher education as to preferred publication processes. Specifically, respondents were questioned on desired standards for authors and editors

in the review process, specific criteria for manuscript selection, authors' rationale for publication, and criteria used in the selection of a journal for manuscript submission.

The results that are most relevant to the present discussion concern the respondents' opinions about the manuscript review process. The issues of most concern dealt with bureaucratic concerns--notification of receipt of manuscript, rapid review process, and publication of policies and operational evaluation systems used in the review process. Interestingly, editors as a group were less likely to agree with the need for multiple reviews of manuscripts, for transmitting critiques to authors, for preserving authors' anonymity, and for having a procedure available for appealing decisions. These results are not surprising when examined from an editor's perspective in that all of these procedures involve extra work for the editor. But they are surprising in that editors define their role from such a limited perspective.

In conclusion it would seem that, especially within the social sciences, journal review processes pose problems for authors. The most commonly mentioned problems include high rejection rate of journals, reservations about the validity of the manuscript evaluation process, especially "particularism," and publication lag times.

The following section describes an evaluation procedure which occurs after the article has been published and which indicates the evaluation of the article by the scientific public.

Citation Analyses. Citation analyses, or counting the frequency with which a particular author or journal article is cited in the scientific literature, have been used as an indicator of the worth of research work

or the researcher. The general assumption of this technique is that the number of citations reflects an article's influence within that scientific field and, therefore, is a measure of its quality. This is especially true since only half of all scientific papers are cited and the average cited paper is referred to only 1.7 times a year (Wade 1975).

Cole and Cole (1971) have suggested three possible uses of citation counts: 1) to distinguish the extent of contributions by various types of scientists; 2) to examine relationships between quality and quantity of output; and 3) to investigate patterns of communication and intellectual linkages within a discipline. In addition, Brittain (1970) has described four uses of citation studies. They are used: 1) to investigate the obsolescence rate of journals, journal articles, and monographs; 2) to investigate the characteristics of citation practices; 3) to study author and journal hierarchies; and 4) to analyze the scattering of literature across time and journals.

The use most relevant to the present discussion is the use of citation counts as an indication of the evaluation of a particular journal article. That is, it is assumed that articles that are cited most frequently in subsequent articles are evaluated more highly than those that are cited infrequently. Lin and Nelson (1969) have differentiated between the use of the term citation and the use of the term reference. Citation refers to each time a reference is cited in a text; reference refers to a work being included in the bibliography, which would occur only once in a given publication. Lin and Nelson compared the results using citation data and reference data and found few differences between the two measures.

Despite the fact that citation analysis may provide a useful,

quantitative measure of scientific quality, its use has not been accepted without objections. Cole and Cole (1971) have delineated nine problems associated with the use of the Science Citation Index, which is the most frequently used source of such data. Some of the problems mentioned by Cole and Cole relate to the use of citation analysis in the evaluation of journal articles and some in the evaluation of individuals and, therefore, they may vary in their relevance to the present discussion. The problems are:

1. Errors in evaluation. It is possible to misclassify a work that is being resisted by the leaders in the field or that has been judged inaccurately (delayed recognition).
2. Critical citations. Citations may refer to papers being criticized and rejected.
3. Treating all citations as equal units. Differentiation should be made between first-rank scientists who cite a work and other citations.
4. Quantity and quality of research output. Although a relationship may exist between quantity and quality of work, it has not been completely substantiated and may vary by discipline.
5. Size of scientific field. The number of citations may be a function of the number of people working in the field, the number of journals published, and the amount of work being published.
6. Contemporaneity of science. One must take into account the dates of publication when comparisons are made between papers because the half-life of papers is short.

7. Integration of basic ideas. Many ideas that are basic in a field (e.g. the law of effect in psychology) are cited without references.
8. Citations to collaborative papers. Such citations are listed in Science Citation Index only by name of first author.
9. Clerical problems. Authors with the same names are not differentiated.

Regardless of these problems, the authors concluded that it is possible to use straight counts with a reasonable degree of confidence. Wade (1975) also made a positive evaluation of the use of citation analysis for evaluating articles and individuals. He cited as advantages of the approach the fact that it described something real, noting that of the fifty most-cited authors, twelve are Nobel laureates; that it may be especially useful as governmental pressure increasingly demands evaluations from granting agencies such as the National Science Foundation and the National Institute of Health; and that in validation studies, citation counts correlate highly with most, if not all, conventional measures of scientific quality.

## RESEARCH FINDINGS

A number of recent studies have empirically examined the review process used by journals. Most of these focus on one or more of three aspects of evaluation--the criteria used to judge articles quality, evaluation errors, and the validity of the evaluations and judgments.

Gottfredson (1978) has differentiated between prescriptive criteria, or idealized behavior plans, and descriptive criteria, which are summaries

of actual behavior. Evaluation errors can be conceptualized as measurement errors in that they reduce the accuracy of judgments about the quality of an article or report. They include lack of interrater reliability and biases such as discrimination and favoritism. Finally, the third issue, the validity of evaluations, concerns the relationships among the various criteria used to judge article quality.

Criteria for Evaluation. Chase (1970) has reviewed the literature dealing with normative criteria for scientific publication. She discussed six norms of the scientific institution presented by Merton (1957) and Barber (1962) in terms of their contributions to the goals of science. The six norms are universalism, organized skepticism, communism, disinterestedness, rationality, and emotional neutrality. Subsequently, she presented data from a survey in which 191 natural and social science faculty members at a single Big 10 school were asked to rate ten criteria in terms of their importance for scientific writing in their discipline. The ten criteria were: 1) originality; 2) logical rigor; 3) compatibility with generally accepted disciplinary ethics; 4) clarity and conciseness of writing style; 5) theoretical significance; 6) mathematical precision; 7) pertinence to current research in the discipline; 8) replicability of research techniques; 9) coverage of significant existing literature; and 10) applicability to "practical" or applied problems in the field.

As pointed out by the author, the results of the survey indicated that respondents considered technical issues as important as informational contributions to the discipline when judging articles. The four most highly rated criteria were: 1) logical rigor; 2) replicability of research techniques; 3) clarity and conciseness of writing style; and 4) originality.

The responses of physical and social scientists were compared to reveal any differences in the importance of the ten criteria. The results indicated that, "Natural scientists placed more emphasis on the qualities of replicability of research techniques, originality, mathematical precision and coverage of the literature, whereas social scientists gave higher ranking to logical rigor, theoretical significance and applied significance" (p. 263). The author attributed these differences to the stages of development in the discipline. Beyer (1978) found reasonable agreement with Chase's findings in her survey of editors of journals in chemistry, physics, sociology, and political science. Smigel and Ross (1970) examined the reviews for Social Problems. The three most frequently mentioned reasons for rejecting a manuscript were: 1) theory or concepts incorrectly or inadequately used; 2) poorly written or presented; and 3) methodology poor or incorrect. The most frequently cited reasons for acceptance were: 1) the paper was interesting; 2) it was significant or meaningful; and 3) it was well written. According to Wolff (1970), the most important criteria used by editors of clinical psychology journals were: 1) contribution to knowledge; 2) research design; and 3) objectivity.

Garvey, Lin, and Nelson (1970) investigated the causes for rejection, comparing social and physical sciences. For both, the most frequently cited reason for manuscript rejection was the inappropriateness of the subject matter for the journal in question. The meaning of the term inappropriate varied for the two areas, however. For physical science inappropriate generally meant submission of applied research to a basic research journal, while for social science inappropriate was in many

cases a euphemism for other reasons for rejection. Larger differences between the meaning of the term inappropriate were exhibited between the two areas, especially involving statistical or methodological problems and theoretical or interpretational grounds.

In another study designed to assess the reliability of ratings made by journal article referees, Scott (1974) utilized seven intuitively derived attributes to evaluate articles. They were: 1) probable interest of readers in the problem; 2) importance of the present contribution to the problem; 3) attention to relevant literature; 4) adequacy of research design and analysis; 5) style and organization of report; 6) succinctness; and 7) recommendation to accept/reject. The author pointed out that the criteria were chosen to be general, descriptive attributes applicable to most journal articles rather than specific criteria which would be applicable to fewer articles. He found that interreferee agreement on these attributes ranged from .07 (probably reader interest in problem) to .37 (attention to relevant literature).

In a somewhat more empirically based study, Gottfredson (1978) investigated evaluative criteria of psychological journal articles. A sample of editors, associate editors, and consulting editors associated with selected journals were asked for their opinions of the relative quality of an article that might be described by each of eighty-three attributes. In order for the number of attributes to be reduced to a set of overall dimensions, the data were subjected to factor analysis. The first five dimensions on which the greatest number of items loaded were defined as follows:

1. A list of "don't" practices to avoid if we want our peers to evaluate our work highly (e.g., "The research was poorly executed," "The author misinterprets the results.")
2. Practices dealing with scientific and substantive matters which are associated with article quality (e.g., "It has excellent generalizability," "It attempts to unify the field.")
3. Practices dealing with stylistic, compositional, or expository matters which are associated with article quality (e.g., "It is well written," "It avoids unrealistic speculation.")
4. The importance of originality (e.g., "It offers a new perspective on an old problem," "It provides new ideas for other investigators.")
5. Triviality of findings or problem addressed (e.g., "The problem addressed is trivial," "The results are trivial or unimportant.")

As a test of the hypothesis that groups in various subdisciplines of psychology differed with respect to their ratings of the desirability of journal articles' characteristics, a discriminant function analysis was performed. The results indicated that although statistically significant in predicting group membership, the resultant discriminant functions were of little practical significance. Using the discriminant functions, it was only possible to correctly predict group membership for 18% of the cases.

These results were interpreted as indicating substantial agreement among the psychologists in their ratings of the desirability of the characteristics and their utilization of the overall dimensions.

Reliability of Judgments. As mentioned previously, evaluation errors represent inaccuracy in the ratings made by judges. Various

articles have addressed the problem of assessing evaluation errors made in the editorial process, and most have focused on the reliability of ratings. Reliability involves the agreement between evaluations made by different judges, between evaluations made at different times, or between different although similar measures. The importance of reliability is that the validity of the evaluations (i.e., the extent to which they reflect article quality) is limited by the reliability of the ratings. If judges cannot agree on the quality or acceptability of a journal article, it can usually be concluded that the evaluations, as measured, do not accurately reflect these attributes.

In the article mentioned previously, Scott (1974) studied the reliability of referees' evaluations of articles submitted to the Journal of Personality and Social Psychology, using a one-page appraisal form which included the seven attributes mentioned above. The appraisal form was included with the manuscript when the latter was sent to various reviewers. In all, double reviews were received for 287 of the manuscripts, and these served as the basis for the results of the study. His results, which are presented in Table 1, indicated that, in general, interreferee agreement was above chance, although not substantially, for six of the seven attributes. Specifically, the highest level of agreement was for the attributes of succinctness and attention to relevant literature while the lowest level of agreement was for probable reader interest and adequacy of research design and analysis. In addition, there was evidence of substantial halo error in that ratings made by individual judges on attributes were highly intercorrelated. The author cautioned that the study was not intended to be a highly controlled investigation but rather to document "two years' experience of one associate editor."

Table 1

Interreferee Agreement on Attributes of Journal Articles

| Attribute | Intraclass Correlation |
|---|---|
| 1. Probable reader interest in problem | .07 |
| 2. Importance of present contribution | .28 |
| 3. Attention to relevant literature | .37 |
| 4. Design and analysis | .19 |
| 5. Style and organization | .25 |
| 6. Succinctness | .31 |
| 7. Recommendation (accept/accept with revisions/reject) | .16 |

Scott (1974) suggested some techniques for increasing interreferee agreement. Lengthening the appraisal form used to evaluate the articles would most likely increase the reliability of the judgments. Another way to increase agreement would be to select pairs of judges with similar perspective on the problem and method. This would, however, reduce the possibility of improving the editorial process by considering differing points of view. Increasing the number of judges would also enhance the reviewing process, although it would necessitate additional time on the part of the pool of judges.

In his article, Gottfredson (1978) reviewed studies of the reliability of peer review processes in psychology and reported correlations from .11 (Bowen, Perloff, and Jacoby 1972) to .84 (McReynolds 1971). Gottfredson concluded that agreement between manuscript rates tends to be low but that agreement about the desirability of specific normative

criteria tends to be somewhat higher. This means that psychologists agree on the criteria to be used in evaluating manuscripts but disagree in their judgments of how well a particular article meets these criteria.

Gottfredson (1978) also differentiated between interjudge reliability and intrajudge reliability. The former refers to agreement across judges with respect to manuscript evaluations, and the latter is a measure of an individual judge's consistency in evaluations as well as a measure of the internal consistency of the evaluation instrument. Utilizing his research on the normative criteria used for manuscript evaluation, Gottfredson developed an evaluation scale. This scale included thirty-six items relating to specific criteria, three items relating to global assessments of quality, and two dealing with the impace of the articles. The last two items dealt with impact on the specific subject matter area and on psychological knowledge.

The results are summarized in Table 2. They indicated that for evaluations of overall quality and impact, internal consistency was quite high but interjudge agreement was relatively modest. For the evaluative scales relating to specific criteria, internal consistency for all but two scales was acceptable. Four of the scales showed rather low interjudge reliability. The author explained both of these results in terms of lack of variance in the subsample for these scales. Gottfredson concluded that his results demonstrated greater reliability of peer judgments of article quality than previous studies had done.

Table 2

Internal Consistency and Interjudge Reliability for Evaluation Scales

| Scales | Internal Consistency [a] | Interjudge Reliability [b] |
|---|---|---|
| 1. Overall quality | .92 | .41 |
| 2. Overall impact | .85 | .35 |
| 3. Don'ts | .78 | .16 |
| 4. Substantive do's | .58 | .50 |
| 5. Stylistic/compositional do's | .74 | .20 |
| 6. Originality | .86 | .37 |
| 7. Trivia | .89 | .40 |
| 8. Where do we go from here? | .64 | .45 |
| 9. Data grinders | .70 | .49 |
| 10. Ho-hum research | .70 | .22 |
| 11. Magnitude of problem | .13 | .19 |

[a] Cronbach's alpha

[b] Intraclass correlation

from Gottfredson (1978)

In reaction to Gottfredson's findings, which they found discouraging, Sears and Weber (1978) presented the results of an assessment of reviewer agreement for manuscripts submitted to the American Psychologist. Manuscripts were rated on a five-point scale with "1" representing "reject" and "5" representing "accept in present form." Of eighty-seven paired ratings, reviewers agreed on fifty-seven, with strong agreement on the categories of "reject" and "reject-resubmit." The authors concluded

that these results indicate fairly high interrater reliability for disposition but not necessarily for reasons for decisions.

Relationships among Criteria. One way to assess the adequancy of evaluations made of journal articles is to study the relationship that exists among the various criteria. To the extent that all of these criteria represent a measure of manuscript "quality," there should be consistent relationships among them.

One of the most extensively studied relationships is that between citation counts and other criteria. These correlations can focus on individual scientists' productivity, in which case citation counts are correlated with such measures as bibliographic counts, quality of graduate education, and peer nomination. Although the majority of investigations have centered on productivity, these are not relevant to the present topic and will not be reviewed. In other cases, investigations of the validity of citation counts emphasize journal article quality. In these studies, citation counts are correlated with experts' judgments of quality and/or impact.

Gottfredson (1978) examined the validity of citation counts by assessing the relationship between them and peer judgments of quality and impact. As mentioned above, quality was judged as a general, overall characteristic as well as being assessed on specific criteria. Results indicated that, although statistically significant, correlations between citation measure and experts' judgments were weak. The largest was .37, between total citations and judgment of impact. All correlations between citation counts and specific criteria were smaller than those between citations and overall evaluations. Interestingly, although Gottfredson

made a correction for self-citations, it appeared that this was not necessary as the correlations for total citations and total citations by others were quite similar.

In addition to assessing the relationship between citation counts and judgments, Gottfredson investigated the correlation among the experts' (individuals nominated by the article authors as competent to evaluate their articles) judgment of quality and impact. Experts gave ratings on a seven-point scale for five criteria: 1) evaluation relative to other works in same time/topic; 2) evaluation relative to other works in any time/topic; 3) overall judgment of scientific quality regardless of subject matter or publication date; 4) impact on specific subject-matter area; and 5) impact on psychological knowledge in general.

Individual correlations are presented in Table 3. Correlations among the three quality measures and between the two impact scales were highest. Because of this, the author summed quality and impact items to produce a "quality scale" and an "impact scale." The correlation between these two was .58 (N=378).

Table 3

Correlations Among Evaluations

| Scale | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. Evaluation relative to other works (same time/topic) | | .84 | .74 | .53 | .48 |
| 2. Evaluation relative to other works (any time/topic) | | | .78 | .48 | .49 |
| 3. Overall quality | | | | .52 | .52 |
| 4. Impact on subject matter | | | | | .74 |
| 5. Impact on psychological knowledge | | | | | |

from Gottfredson (1978)

## CONCLUSIONS

Based on the material reviewed above, a few conclusions can be drawn regarding the process and results of evaluation of research reports and journal articles. First, the strengths and weaknesses of the various evaluation processes complement one another. The informal evaluation that takes place prior to the submission of the paper for publication is quite subjective, but provides valuable feedback to the author. It has been shown that based on peer evaluation, authors often revise their papers in terms of research analysis, review of the literature, or presentation of findings.

The editorial process, on the other hand, is somewhat more objective and usually includes less descriptive information on the specific strengths and weaknesses of the paper than is the case with the informal peer evaluation. It does, however, usually provide feedback to the author, along with the editorial decision. The editorial process primarily serves a gate-keeping function. That is, it evaluates and screens the number of articles published in journals, thus helping to ensure the quality of the information provided to the user. It also reduces the need for the user to sift through numerous documents, evaluating them in terms of quality and relevance.

Citation counts are perhaps the most objective indicator of the evaluation of a journal article and have been shown to have validity as measured by their relationship with other indicants of article quality. Citation counts do not, however, communicate much specific information as to the sources of article quality (e.g., rigor of research design, review of the literature).

The research that has been conducted to investigate the adequacy of the evaluations that are made and the possibility of improving this process indicates a fairly strong agreement among judges on normative criteria, or what constitutes quality in a journal article. However, this agreement does not carry over to evaluations made as to the quality of actual articles. This lack of reliability most probably is a function of a variety of factors. Among those cited have been lack of a standardized process and/or rating form and a lack of agreement as to the relative importance of normative criteria. Especially in the social sciences, judges may be emphasizing various criteria differentially (e.g., heuristics vs. rigor of research design). Most relevant to the social sciences is possibly the lack of a paradigm that would facilitate the evaluation process. For example, in the social sciences there is disagreement on what is a good theory or what are appropriate methods; therefore, "particularism" influences the decision-making process.

Another conclusion relating to the editorial process concerns whether the rater should be aware of the author and the author's institutional affiliation while reviewing the article. Although not completely conclusive, most research shows that the judgments of raters can be affected by personal characteristics of the authors (e.g., where they did their doctoral work, their sex, Moore 1978). Therefore, it would seem reasonable to make a practice of deleting the name and institutional affiliation of the author from the manuscript during the review process.

Rodman and Mancini (1977) pointed out three potential biases in refereeing that have not been examined. These are sponsored submissions where a friend or mentor of the author endorses the manuscript; inside

track where the author and the editor have a special relationship (e.g., both are members of the same institution or the author is a member of the editorial board, etc.), and back region where reviewer makes comments for the editor's eyes only so that the author cannot refute them.

Among the recommendations made which might improve the evaluation process are increasing the number of reviewers and using a standardized form to evaluate manuscripts on specific criteria. Because implementation of these recommendations may be resisted by reviewers and editors, additional research is necessary to determine the acceptability of various modifications in the review process. Additional research is also needed to clarify the nature of an adequate paradigm in the social sciences and the relative importance of normative criteria.

## REFERENCES


Baker, D.B. Communication or chaos? Science, 1970, 169, 739-742.

Balaban, M., ed. Scientific information transfer: The editor's role. Dordrecht, Holland: D. keidel Publishers Co., 1978.

Barber, B. Science and social order. New York: Cowell Collier and Macmillan, 1962.

Beyer, J. M. Editorial policies and practices among leading journals in four scientific fields. Sociological Quarterly, 1978, 17, 66-88.

Bowen, D. D.; Perloff, R.; and Jacoby, J. Improving manuscript evaluation procedures. American Psychologist, 1972, 27, 221-225.

Brackbill, Y., and Korten, F. Journal reviewing practices: Authors' and APA members' suggestions for revision. American Psychologist, 1970, 25, 937-940.

Brittain, J. M. Information and its uses. New York: John Wiley & Sons, Inc., 1970.

Chase, J. Normative criteria for scientific publication. American Sociologist, 1970, 8, 187-189.

Cole, J., and Cole, S. Measuring the quality of sociological research: Problems in the use of the Science Citation Index. American Sociologist, 1971, 6, 23-29.

Crane, D. Social structure in a group of scientists: A test of the invisible college hypothesis. American Sociological Review, 1969, 34, 335-352.

Garvey, W. D., and Gottfredson, S. D. Changing the system: Innovations in the interactive system of scientific communication. Journal of Information Processing and Management, 1976, 12, 165-176.

Garvey, W. D.; Lin, N.; and Nelson, C.E. Communication in the physical and social sciences: The process of disseminating and assimilating information differs in these two groups of sciences. Science, 1970, 170, 1166-1173.

Garvey, W. D.; Lin, N.; and Tomita, K. Research studies in patterns of scientific communication: Information-exchange processes associated with the production of journal articles. Information Storage and Retrieval, 1972, 8, 207-221.

Gottfredson, S. D. Evaluating psychological research reports: Dimensions, reliability, and correlating or quality judgments. American Psychologist, 1978, 33, 920-934.

Kuhn, T. S. The structure of scientific revolutions. Chicago: The University of Chicago Press, 1970.

Lin, N., and Nelson, C. E. Bibliographic reference patterns in some sociological journals, 1965-1966. American Sociologist, 1969, 4, 47-50.

Lindsey, D. The scientific publication system in social science. San Francisco: Jossey-Bass Publishers, 1978.

Mahoney, M. J. Publication prejudices: An experimental study of confirmatory bias in the peer review system. Cognitive Therapy and Research, 1977, 1, 161-175.

Manheim, F. T. Referees and the publication crisis. Transactions, American Geophysical Union, 1973, 54, 532-537.

McReynolds, P. Reliability of ratings of research papers. American Psychologist, 1971, 26, 400-401.

Merton, R. K. Social theory and social structure. Glencoe, Illinois: Free Press, 1957.

Moore, M. Discrimination or favoritism? Sex bias in book reviews. American Psychologist, 1978, 33, 936-938.

Nelson, C. E. The communication system surrounding archival journals in educational research. Educational Researcher, 1972, 1, 13-15.

Newman, S. H. Improving the evaluation of submitted manuscripts. American Psychologist, 1966, 21, 980-981.

Pfeffer, J.; Leong, A.; and Strehl, K. Paradigm development and particularism: Journal publication in three scientific disciplines. Social Forces, 1977, 55, 938-951.

Price, D. J. de S. Big science, little science. New Haven, Connecticut: Yale University Press, 1964.

Rodman, H., and Mancini, J. A. Editors, manuscripts, and equal treatment. Research in Higher Education, 1977, 7, 369-374.

Scott, W. A. Interreferee agreement on some characteristics of manuscripts submitted to the Journal of Personality and Social Psychology. American Psychologist, 1974, 29, 698-702.

Sears, S., and Weber, B. L. R. The reliability of reviews for the American Psychologist. American Psychologist, 1978 33, 935.

Silverman, R. J., and Collins, E. L. Publishing relationships in higher education. Research in Higher Education, 1975, 3, 365-382.

Smigel, E., and Ross, H. L. Factors in editorial decision. American Sociologist, 1970, 5, 19-21.

Storer, N. W. The social system of science. New York: Holt, Rinehart and Winston, 1966.

Wade, N. Citation analysis: A new tool for science administrators. Science, 1975, 188, 429-432.

Wolff, W. M. A study of criteria for journal manuscripts. American Psychologist, 1970, 25, 636-639.

Yoels, W. C. The structure of scientific fields and the allocation of editorships on scientific journals: Some observations on the politics of knowledge. Sociological Quarterly, 1974, 15, 264-276.

Ziman, J. The light of knowledge. ASLIB Proceedings, 1970, 22, 186-200.

Zuckerman, H., and Merton, R. K. Patterns of evaluation in science: Institutionalism, structure and functions of the referee system. Minerva, 1971, 9, 66-100.

ERIC/TM Report 77

THE REVIEWING PROCESS IN SOCIAL SCIENCE PUBLICATIONS:

A REVIEW OF RESEARCH

by

Susan E. Hensley and Carnot E. Nelson

Because of the increase in the number of manuscripts submitted to journals, editors and publishers must either publish more manuscripts or increase the proportion of manuscripts rejected. Rejection decisions are usually informed by informal peer evaluation, editorial processes, and/or citation analysis. Unfortunately, no standard criteria for evaluation have been developed, and interreferee agreement on the attributes of journal articles is generally low. The research on these issues is reviewed.

---

ORDER FORM

Please send _____ copies of ERIC/TM Report 77, "The Reviewing Process in Social Science Publications," at $4.00 each.

Name_____

Address_____

_____ Zip_____

Return this form to:

ERIC/TM

Educational Testing Service

Princeton, NJ 08541

RECENT TITLES

IN THE ERIC/TM REPORT SERIES

#76 - Intelligence Testing, Education, and Chicanos: An Essay in Social Inequality, by Adalberto Aguirre, Jr. 40p. $5.50.

#75 - Contract Grading, by Hugh Taylor. 95p. $7.50.

#74 - Intelligence, Intelligence Testing, and School Practices, by Richard De Lisi. 32p. $4.50.

#73 - Measuring Attitudes Toward Reading, by Ira Epstein. 120p. $9.50.

#72 - Methods of Identifying Gifted Minority Students, by Ernest M. Bernal. 32p. $4.50.

#71 - Sex Bias in Testing: An Annotated Bibliography, by Barbara Hunt, 85p. $5.00.

#70 - The Role of Measurement in the Process of Instruction, by Jeffrey K. Smith, 29p. $3.50.

#69 - Client Assessment: A Manual For Employment and Training Agencies, by Thomas E. Backer.

Volume I: Introduction to Assessment Program Development 50p. $4.00.

Volume II: Assessment Techniques. 49p. $4.00. ($7.00 for both volumes).